# RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _10/550,155_

Source: _1FWP_

Date Processed by STIC: _11/15/06_

# ENTERED

RAW SEQUENCE LISTING                    DATE: 11/15/2006
PATENT APPLICATION:  US/10/550,155      TIME: 11:52:04

Input Set : F:\56446-20040.40 SEQ (client).txt
Output Set: N:\CRF4\11152006\J550155.raw

```
 4 <110> APPLICANT: Diversa Corporation
 5        K. Gray
 6        J. Garrett
 7        N. Aboushadi
 8        R. Knowles
 9        E. O'Donoghue
10        E. Waters
12 <120> TITLE OF INVENTION: GLUCOSIDASES, NUCLEIC ACIDS ENCODING
13        THEM AND METHODS FOR MAKING AND USING THEM
15 <130> FILE REFERENCE: 564462004040
```
C--> 17 <140> CURRENT APPLICATION NUMBER: US/10/550,155
C--> 17 <141> CURRENT FILING DATE: 2005-09-20
```
17 <150> PRIOR APPLICATION NUMBER: 60/456,972
18 <151> PRIOR FILING DATE: 2003-03-20
20 <160> NUMBER OF SEQ ID NOS: 24
22 <170> SOFTWARE: FastSEQ for Windows Version 4.0
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 1710
26 <212> TYPE: DNA
27 <213> ORGANISM: Bacteria
29 <400> SEQUENCE: 1
30 atgcagcgcc atcgcagaag gtgcagggct aagcttgtcg ggttcgtttt ggcaccccgt    60
31 ttggcggggtg catggaagcc cggaggggggg ccctcgatgt cgcagactcc atggtggcgc   120
32 ggtgccgtca tctaccagat ctacccgcgc agtttcctcg acgccaacgg cgacggggtc   180
33 ggtgacctgc ccggcatcat cgaccggctg gagtacgtgg ccgcgctggg cgtggacgcc   240
34 atctgggtct cgccgttctt cacctcgccg atggccgatt tcggctacga catcgccgac   300
35 catcgcgacg tggacccgct gtttggcacg ctggccgatt tcgaccggct gctggccaag   360
36 gcgcatgcgc tgggcctgaa ggtgatgatc gaccaggtgt tcagccacac ctcgatcgac   420
37 cacgcctggt tccgtgagag ccggcaggac cgcaccaatc cgaaggcgga ctggtacgtg   480
38 tgggccgacc cgcgcgagga cggcacgccg cccaacaact ggatgtcgat cttcggcggg   540
39 gtggcctggc aatgggagcc gcgccgggag cagtacttcc tgcacaactt cctggccgac   600
40 cagccggacc tggatttcca caacccggcg gtgcagcagg ccacgctgga ctacgtgcgc   660
41 ttctggctgg accggggcgt ggacgggttc cgcctggact cgatcaactt ctgcttccac   720
42 gacgcgcagt tgcgcgacaa cccggccaag ccgctggaaa agcgcgtcgg ccgtggcttc   780
43 agcgcggaca atccgtacgc ctaccagtac cactactaca acaacaccca gccggagaac   840
44 atcggcttca tcgagcgcct gcgtggggttg ctggacgaat acccgggcac cgtgagcctg   900
45 ggcgagatct cggccgagga ctcgctggcc accaccgccg agtacaccgc gccgggccgc   960
46 ctgcacatgg gctacagctt cgagctgctg gtgaaggatt tcagcgccgg ctacatccgc  1020
47 gacaccgtgt gcgcggctgga agcgacgatg accgaaggct ggccgtgctg ggcgatctcc  1080
48 aaccacgacg tggagcgtgc ggtcactcgc tggggcggcc atccggcccg ccgcgggctg  1140
49 gcgcggatgc tggtggcgct gctgtgctcg ctgcgtggct cgatctgcct gtaccagggc  1200
50 gaggagctgg gcctgggcga ggcggacgtg ccgttcgagg cgctgcagga cccgtatggc  1260
51 atcaccttct ggccgaactt caagggccgc gacggctgcc gcacgccgat gccgtggatc  1320
```

RAW SEQUENCE LISTING                    DATE: 11/15/2006
PATENT APPLICATION:  US/10/550,155      TIME: 11:52:04

Input Set : F:\56446-20040.40 SEQ (client).txt
Output Set: N:\CRF4\11152006\J550155.raw

```
52 gatgcgccgt tggcagggtt caccagcggt gagccgtggc tgccgattcc ggccgagcac      1380
53 cgcgccgcgg cggtggcggt gcaggagcac gacccgcact cggtgttgaa cgcgttccgc      1440
54 cagttcctgg catggcgcag gacgatgccg acgctgctgg tgggcgacat cgtcttcctg      1500
55 cagacggccg agccggtgct gatgttcgag cgccggcatg cggggggagac gctgctgctg     1560
56 gccttcaacc tggcggccga caccgcgcgc gtggcgctgc ccgccggcag ctggcagccg      1620
57 atgcacgtgc cgggcccgga cgtgggccag gccgacggcg ggacgttggt actgccggcg      1680
58 cagtcgatgt actgcgcgcg cctgggctga                                       1710
60 <210> SEQ ID NO: 2
61 <211> LENGTH: 569
62 <212> TYPE: PRT
63 <213> ORGANISM: Bacteria
65 <220> FEATURE:
66 <221> NAME/KEY: SIGNAL
67 <222> LOCATION: (1)...(24)
69 <221> NAME/KEY: DOMAIN
70 <222> LOCATION: (46)...(434)
71 <223> OTHER INFORMATION: Alpha amylase, catalytic domain
```
W-->  73 <400> 2
```
74 Met Gln Arg His Arg Arg Arg Cys Arg Ala Lys Leu Val Gly Phe Val
75  1               5                   10                  15
76 Leu Ala Pro Arg Leu Ala Gly Ala Trp Lys Pro Gly Gly Gly Pro Ser
77             20                  25                  30
78 Met Ser Gln Thr Pro Trp Trp Arg Gly Ala Val Ile Tyr Gln Ile Tyr
79         35                  40                  45
80 Pro Arg Ser Phe Leu Asp Ala Asn Gly Asp Gly Val Gly Asp Leu Pro
81     50                  55                  60
82 Gly Ile Ile Asp Arg Leu Glu Tyr Val Ala Ala Leu Gly Val Asp Ala
83 65                  70                  75                  80
84 Ile Trp Val Ser Pro Phe Phe Thr Ser Pro Met Ala Asp Phe Gly Tyr
85                 85                  90                  95
86 Asp Ile Ala Asp His Arg Asp Val Asp Pro Leu Phe Gly Thr Leu Ala
87             100                 105                 110
88 Asp Phe Asp Arg Leu Leu Ala Lys Ala His Ala Leu Gly Leu Lys Val
89         115                 120                 125
90 Met Ile Asp Gln Val Phe Ser His Thr Ser Ile Asp His Ala Trp Phe
91     130                 135                 140
92 Arg Glu Ser Arg Gln Asp Arg Thr Asn Pro Lys Ala Asp Trp Tyr Val
93 145                 150                 155                 160
94 Trp Ala Asp Pro Arg Glu Asp Gly Thr Pro Pro Asn Asn Trp Met Ser
95                 165                 170                 175
96 Ile Phe Gly Gly Val Ala Trp Gln Trp Glu Pro Arg Arg Glu Gln Tyr
97             180                 185                 190
98 Phe Leu His Asn Phe Leu Ala Asp Gln Pro Asp Leu Asp Phe His Asn
99         195                 200                 205
100 Pro Ala Val Gln Gln Ala Thr Leu Asp Tyr Val Arg Phe Trp Leu Asp
101     210                 215                 220
102 Arg Gly Val Asp Gly Phe Arg Leu Asp Ser Ile Asn Phe Cys Phe His
103 225                 230                 235                 240
104 Asp Ala Gln Leu Arg Asp Asn Pro Ala Lys Pro Leu Glu Lys Arg Val
```

```
105                  245              250                255
106 Gly Arg Gly Phe Ser Ala Asp Asn Pro Tyr Ala Tyr Gln Tyr His Tyr
107          260              265              270
108 Tyr Asn Asn Thr Gln Pro Glu Asn Ile Gly Phe Ile Glu Arg Leu Arg
109       275              280              285
110 Gly Leu Leu Asp Glu Tyr Pro Gly Thr Val Ser Leu Gly Glu Ile Ser
111    290              295              300
112 Ala Glu Asp Ser Leu Ala Thr Thr Ala Glu Tyr Thr Ala Pro Gly Arg
113 305              310              315              320
114 Leu His Met Gly Tyr Ser Phe Glu Leu Leu Val Lys Asp Phe Ser Ala
115          325              330              335
116 Gly Tyr Ile Arg Asp Thr Val Ser Arg Leu Glu Ala Thr Met Thr Glu
117          340              345              350
118 Gly Trp Pro Cys Trp Ala Ile Ser Asn His Asp Val Glu Arg Ala Val
119       355              360              365
120 Thr Arg Trp Gly Gly His Pro Ala Arg Pro Arg Leu Ala Arg Met Leu
121    370              375              380
122 Val Ala Leu Leu Cys Ser Leu Arg Gly Ser Ile Cys Leu Tyr Gln Gly
123 385              390              395              400
124 Glu Glu Leu Gly Leu Gly Glu Ala Asp Val Pro Phe Glu Ala Leu Gln
125          405              410              415
126 Asp Pro Tyr Gly Ile Thr Phe Trp Pro Asn Phe Lys Gly Arg Asp Gly
127          420              425              430
128 Cys Arg Thr Pro Met Pro Trp Ile Asp Ala Pro Leu Ala Gly Phe Thr
129    ·435              440              445
130 Ser Gly Glu Pro Trp Leu Pro Ile Pro Ala Glu His Arg Ala Ala Ala
131    450              455              460
132 Val Ala Val Gln Glu His Asp Pro His Ser Val Leu Asn Ala Phe Arg
133 465              470              475              480
134 Gln Phe Leu Ala Trp Arg Arg Thr Met Pro Thr Leu Leu Val Gly Asp
135          485              490              495
136 Ile Val Phe Leu Gln Thr Ala Glu Pro Val Leu Met Phe Glu Arg Arg
137          500              505              510
138 His Ala Gly Glu Thr Leu Leu Leu Ala Phe Asn Leu Ala Ala Asp Thr
139          515              520              525
140 Ala Arg Val Ala Leu Pro Ala Gly Ser Trp Gln Pro Met His Val Pro
141    530              535              540
142 Gly Pro Asp Val Gly Gln Ala Asp Gly Gly Thr Leu Val Leu Pro Ala
143 545              550              555              560
144 Gln Ser Met Tyr Cys Ala Arg Leu Gly
145          565
147 <210> SEQ ID NO: 3
148 <211> LENGTH: 1293
149 <212> TYPE: DNA
150 <213> ORGANISM: Unknown
152 <220> FEATURE:
153 <223> OTHER INFORMATION: Obtained from an environmental sample
155 <400> SEQUENCE: 3
156 atgagtctgt ggcgtgcgct cgtcgcgttt gttctgctgg ccgtcgccat gcccgcattt          60
```

RAW SEQUENCE LISTING                    DATE: 11/15/2006
PATENT APPLICATION:   US/10/550,155     TIME: 11:52:04

Input Set : **F:\56446-20040.40 SEQ (client).txt**
Output Set: **N:\CRF4\11152006\J550155.raw**

```
157 gccgacgtgg tcgcgaccgc ctcttcgcca ggcgatgtcc tcaaggtgga gatcaccacc   120
158 aacggcgagg gccgcatcgg ctatgcggtc acccggctcg gcaagccggt aatcggcgag   180
159 agccacctcg gattcctcct ggccgacgcg ccgcagctgc tgcgcaactt ccaggtcgtc   240
160 gatcaggcca cccggacctt cgacgaaacg tgggagcagc cgtgggggga gtggcgcacg   300
161 gtccgcaacc actacaacga gctcgcgatc accttcgagg agaagaccaa gctccatcgg   360
162 cggatgcgga tcgttttttcg cctgttcgat gaagggatcg gctttcgcta cgagcttccc   420
163 cggcagccga acctggcgca cgccaacatc gccgaggaac tgacccagtt caacgtcgcg   480
164 cgaccgggca cggcctggtg ggcaccggcg ttcgaatcca accgcgagga atatctctac   540
165 aaccagaccc cgatcgacgg tgtcgcgatt gcgatgactc cattcacgat gcggttcgag   600
166 gacgggactc acctcagcat ccacgaagcc gcgctggtcg actattccgg gatgaacgtc   660
167 acgcgtgtcc agggcacgaa cttcaaggcc atcctgacgc ccggttcgat gggcccccaaa  720
168 gtctcccgcg atacccccatt cgagaccccg tggcgggtca tcctgatcag ccccgacgct   780
169 gcgcacctct acgaatcgaa caggctgatc ctcaacctca acgaacccaa caagctcggc   840
170 gacgtcagct gggtccaccc gcgcaaatat gtcggcatct ggtggggcat gcacctcgat   900
171 acccagagtt gggcctcggg gccgaagcac ggcgcgacca ccgcttatgc gaagcgaatg   960
172 atcgatttcg cggcgacgaa cggctttacc gggctgctcg tcgaaggctg gaacaaggga  1020
173 tgggacggag actggttcgc gaccggcgac gatttcagct tcaccgaacc ctatcccgat  1080
174 ttcgacatcc gagccgtcgc ggcctacagc ctcaggaaag gcgtccacct catcgggcat  1140
175 cacgaaacca gcggcaacat cgcccactac gagcaacagc tcgatgcggc gctcgatctc  1200
176 gaccgccagc tcggcatcga cacggtgaag acgggctacg tctcggatgc cggcggcatc  1260
177 caggcgctcg ggcccgacgg caggatccaa aga                                1293
179 <210> SEQ ID NO: 4
180 <211> LENGTH: 431
181 <212> TYPE: PRT
182 <213> ORGANISM: Unknown
184 <220> FEATURE:
185 <223> OTHER INFORMATION: Obtained from an environmental sample
```
W--> 187 <221> **NAME/KEY: SIGNAL**
```
188 <222> LOCATION: (1)...(21)
```
W--> 190 **<400> 4**
```
191 Met Ser Leu Trp Arg Ala Leu Val Ala Phe Val Leu Leu Ala Val Ala
192 1               5                   10                  15
193 Met Pro Ala Phe Ala Asp Val Val Ala Thr Ala Ser Ser Pro Gly Asp
194             20                  25                  30
195 Val Leu Lys Val Glu Ile Thr Thr Asn Gly Glu Gly Arg Ile Gly Tyr
196         35                  40                  45
197 Ala Val Thr Arg Leu Gly Lys Pro Val Ile Gly Glu Ser His Leu Gly
198     50                  55                  60
199 Phe Leu Leu Ala Asp Ala Pro Gln Leu Leu Arg Asn Phe Gln Val Val
200 65                  70                  75                  80
201 Asp Gln Ala Thr Arg Thr Phe Asp Glu Thr Trp Glu Gln Pro Trp Gly
202             85                  90                  95
203 Glu Trp Arg Thr Val Arg Asn His Tyr Asn Glu Leu Ala Ile Thr Phe
204             100                 105                 110
205 Glu Glu Lys Thr Lys Leu His Arg Arg Met Arg Ile Val Phe Arg Leu
206         115                 120                 125
207 Phe Asp Glu Gly Ile Gly Phe Arg Tyr Glu Leu Pro Arg Gln Pro Asn
208     130                 135                 140
209 Leu Ala His Ala Asn Ile Ala Glu Glu Leu Thr Gln Phe Asn Val Ala
```

```
210 145                     150                     155                     160
211 Arg Pro Gly Thr Ala Trp Trp Ala Pro Ala Phe Glu Ser Asn Arg Glu
212                     165                     170                     175
213 Glu Tyr Leu Tyr Asn Gln Thr Pro Ile Asp Gly Val Ala Ile Ala Met
214                     180                     185                     190
215 Thr Pro Phe Thr Met Arg Phe Glu Asp Gly Thr His Leu Ser Ile His
216             195                     200                     205
217 Glu Ala Ala Leu Val Asp Tyr Ser Gly Met Asn Val Thr Arg Val Gln
218         210                     215                     220
219 Gly Thr Asn Phe Lys Ala Ile Leu Thr Pro Gly Ser Met Gly Pro Lys
220 225                     230                     235                     240
221 Val Ser Arg Asp Thr Pro Phe Glu Thr Pro Trp Arg Val Ile Leu Ile
222                     245                     250                     255
223 Ser Pro Asp Ala Ala His Leu Tyr Glu Ser Asn Arg Leu Ile Leu Asn
224                     260                     265                     270
225 Leu Asn Glu Pro Asn Lys Leu Gly Asp Val Ser Trp Val His Pro Arg
226             275                     280                     285
227 Lys Tyr Val Gly Ile Trp Trp Gly Met His Leu Asp Thr Gln Ser Trp
228         290                     295                     300
229 Ala Ser Gly Pro Lys His Gly Ala Thr Thr Ala Tyr Ala Lys Arg Met
230 305                     310                     315                     320
231 Ile Asp Phe Ala Ala Thr Asn Gly Phe Thr Gly Leu Leu Val Glu Gly
232                     325                     330                     335
233 Trp Asn Lys Gly Trp Asp Gly Asp Trp Phe Ala Thr Gly Asp Asp Phe
234             340                     345                     350
235 Ser Phe Thr Glu Pro Tyr Pro Asp Phe Asp Ile Arg Ala Val Ala Ala
236         355                     360                     365
237 Tyr Ser Leu Arg Lys Gly Val His Leu Ile Gly His His Glu Thr Ser
238     370                     375                     380
239 Gly Asn Ile Ala His Tyr Glu Gln Gln Leu Asp Ala Ala Leu Asp Leu
240 385                     390                     395                     400
241 Asp Arg Gln Leu Gly Ile Asp Thr Val Lys Thr Gly Tyr Val Ser Asp
242                     405                     410                     415
243 Ala Gly Gly Ile Gln Ala Leu Gly Pro Asp Gly Arg Ile Gln Arg
244             420                     425                     430
246 <210> SEQ ID NO: 5
247 <211> LENGTH: 1773
248 <212> TYPE: DNA
249 <213> ORGANISM: Bacteria
251 <400> SEQUENCE: 5
252 atgaccctca acaacaccca tgccgactgg tggaaacagg cggtggtcta ccaggtctac       60
253 ccgcgcagct tcaaggatgt gaacggtgac ggtctgggcg acatcgccgg cgtgaccgac      120
254 cgcatcccct atctcaagga gctgggcgtc gacgcgatct ggttgtcgcc gttctacccg      180
255 tccgagctgg cggacggcgg ctacgacgtc atcgactacc gcgacgtgga cccgcgcctg      240
256 ggctccatgg acgacttcga cgccatggcc gcagccgcgc acgaggccgg catgaaggtg      300
257 atcgtggaca tcgtgccgaa ccacacctcc gaccggcacg tctggtttga ggaggcgctg      360
258 gcggcagaac cgggctcccc cgcgcgcgac cggtacatct tccgcgacgg tctgggcgaa      420
259 cacggcgagc tgcccccgaa cgactggcag tcgatcttcg gcggcggcgc atgggaacgg      480
260 gtgccggacg gccagtggta cctgcacatg ttcgccaagg aacagcccga cctcaactgg      540
```

**VERIFICATION SUMMARY**        DATE: 11/15/2006
PATENT APPLICATION:  US/10/550,155  .  TIME: 11:52:06

Input Set : **F:\56446-20040.40 SEQ (client).txt**
Output Set: **N:\CRF4\11152006\J550155.raw**

```
L:17 M:270 C: Current Application Number differs, Replaced Current Application No
L:17 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:73 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:2
L:187 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:190 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:4
L:562 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:566 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:10
L:683 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:687 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:12
L:803 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:807 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:14
L:913 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:917 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:16
L:1029 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:1033 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:18
L:1142 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:1146 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:20
L:1371 M:281 W: Numeric Fields not Ordered, <221> Sort in ascending order!
L:1375 M:258 W: Mandatory Feature missing, <220> Tag not found for SEQ ID#:24
```